Intrinsically Linear Regression

Chapter 9

Introduction

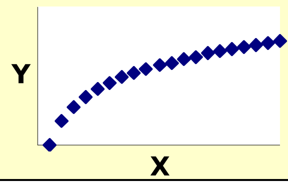
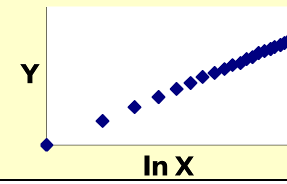
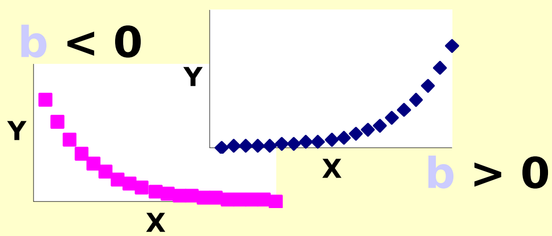
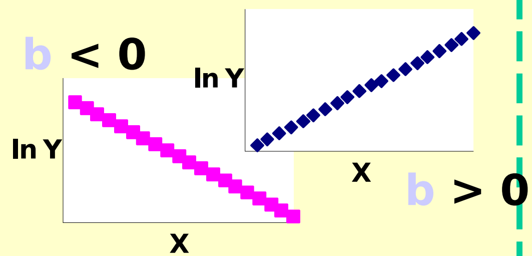
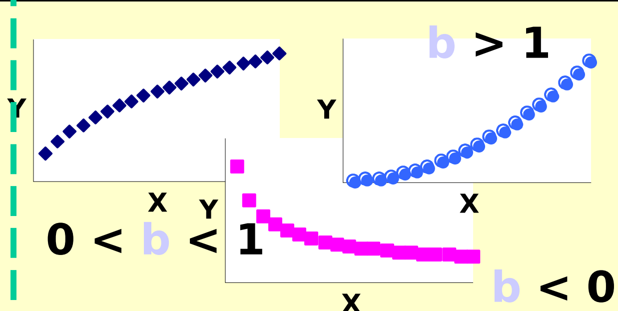
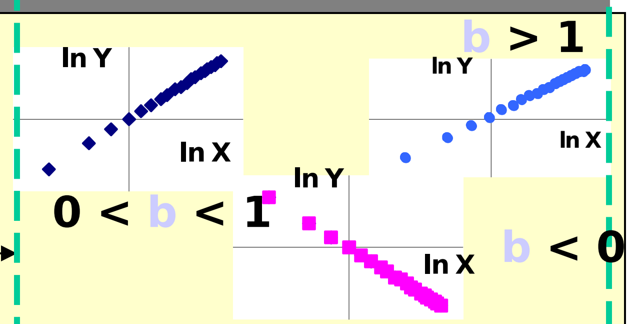
- **In Chapter 7 we discussed some deviations from the assumptions of the regression model.**
- **One of the assumptions was that the residuals are normally distributed.**
- **If this assumption does not hold, then we may have to transform the data into a form that will make it appear linear, so that regression analysis can be used.**

Non-linear Models – Introduction

- To model non-linear relationships with OLS regression, the data must first be transformed in a way that makes the relationship linear
- All the steps for linear regression may then be performed on the transformed data
- The most common forms of non-linear models are:
  - Logarithmic
  - Exponential
  - Power

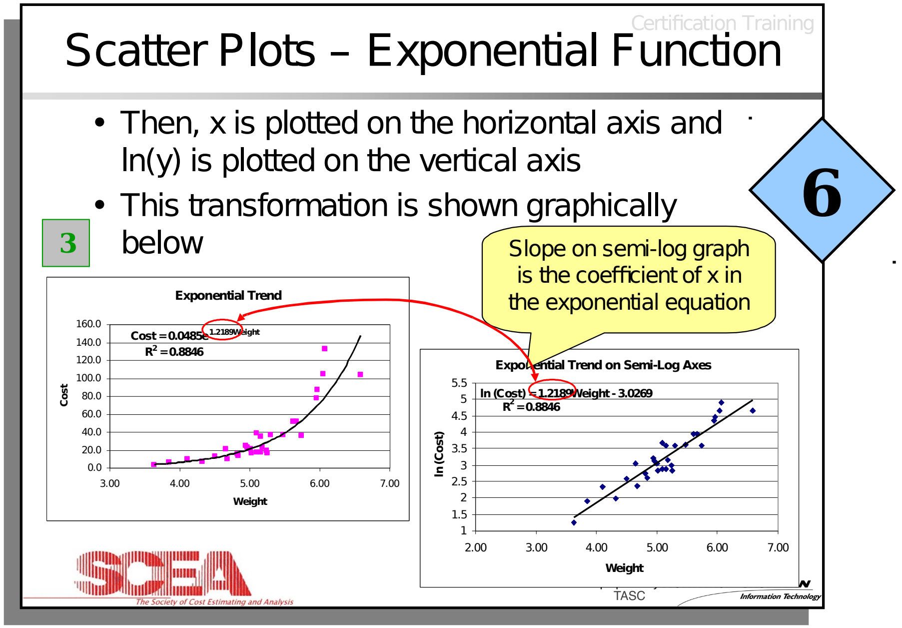
Linear transformations

12

Unit Space	Model	Log Space
	Logarithmic $y = a + b \ln x$	
	Exponential $y = a\, e^{b x}$ $\ln y = \ln a + b\, x$	
	Power $y = a\, x^{b}$ $\ln y = \ln a + b \ln x$	

Example: Exponential Model

- The data is plotted in unit space (left) then trans-formed and plotted on a semi-log graph (right)
- The next step is to conduct linear regression analysis on the data in semi-log space
- After the analysis is complete, we will transform the parameters of the linear equation back to unit space



$$a = e^{\ln a}$$

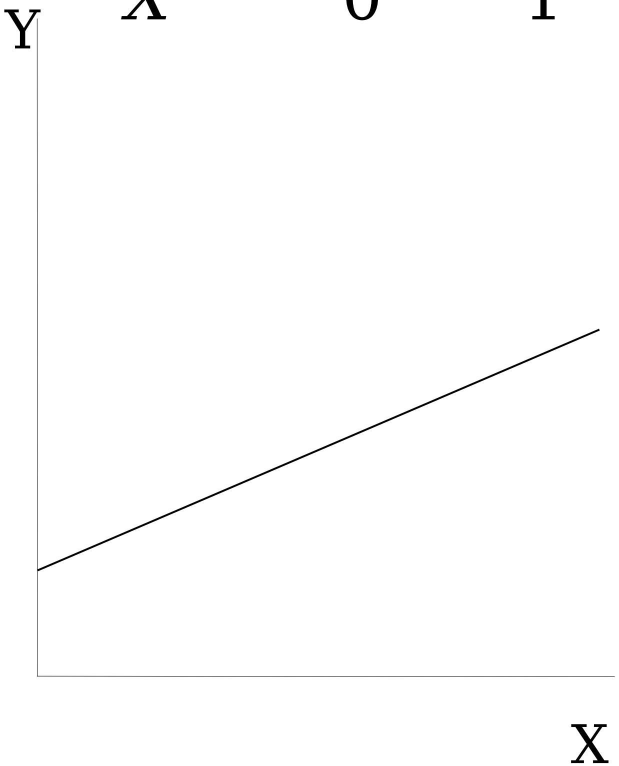
$$\ln y = \ln a + b\,x \iff y = a\,e^{b\,x}$$

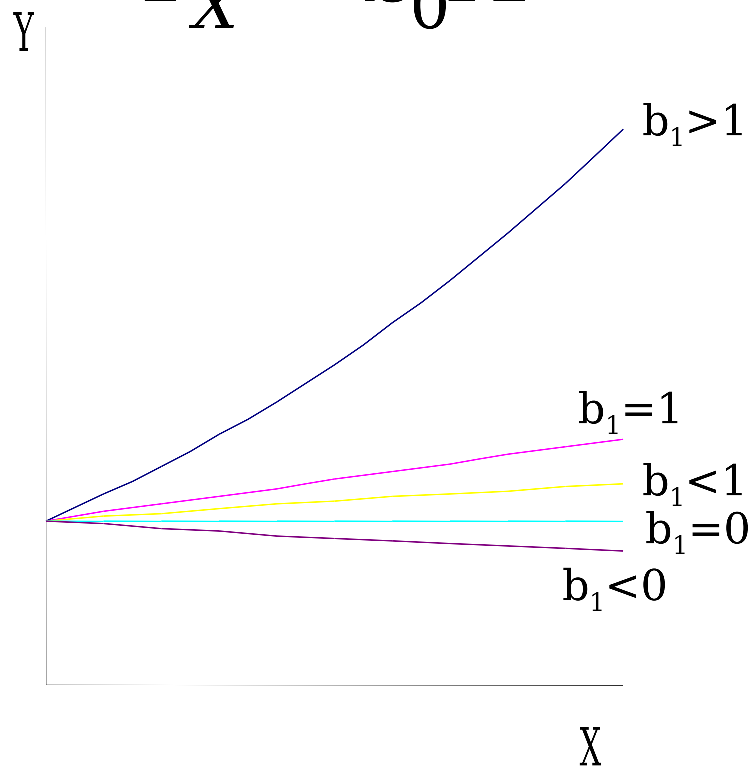
$$b = b$$

The Multiplicative Model

$$\hat{Y}_X = b_0 + b_1 X$$

$$\hat{Y}_X = b_0 X^{b_1}$$





- **Linear Equation: A unit change in X causes Y to change by b_1**
- **Multiplicative Equation: A change in X causes Y to change by a percentage proportional to the change in X**

The Multiplicative Model

- However, in order to produce a cost estimating relationship using the method of least squares, we must transform the multiplicative model into a linear model (at least temporarily).
- The solution is to create a log-linear equation.

$$\hat{Y} = AX^{b_1} \quad \Leftrightarrow \quad \ln(\hat{Y}) = b_0 + b_1 \ln(X)$$

- Now we can perform a linear regression on ln(Y) and ln(X), then transform the results of the linear regression into an exponential equation.

Interpreting the Results

- **A linear regression of transformed data will provide exactly the same *type* of results as a linear regression of raw data.**
  - **The computer doesn't know you've given it transformed data.**
- **So you need to re-transform the results into an exponential model.**
- **The Intercept corresponds to ln(b_0), and the slope corresponds to the exponent, b_1.**

$$A = e^{Intercept} = e^{b_0}$$

$$b_1 = b_1$$

- **Moreover, the statistics of the transformed data can be misleading.**

Interpreting the Results

- **The Standard Error and the R^2 reported for a log-linear model can not be compared to those for a linear model. This is because both are functions of SSE.**
- **Recall that SSE is the error sum of squares, and the standard error is expressed in terms of dollars.**

$$SE_{unit} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n-k-1}} = \$XXX$$

- **The Standard Error in log space has a different meaning than that in unit space.**

$$SE_{\log} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum (\ln(Y_i) - \ln(\hat{Y}))^2}{n-k-1}} = X.XX$$

- **However, we *can* compare between log-linear models.**

Interpreting the Results

- **When comparing linear and log-linear cost models, use section III in Cost$tat entitled Predictive Measures (unit space).**
- **Cost$tat provides measures in unit space in section III for log-linear models. These numbers can be compared directly to corresponding linear models.**
  - **Standard Error**
  - **Coefficient of Variation**
  - **Adjusted R-squared**
- **Remember, you CANNOT compare these measures in log space to the same measures in unit space. All comparisons must be in unit space.**